S. TURNER

RAW SEQUENCE LISTING
PATENT APPLICATION **US/09/235,416A**

DATE: 04/26/2000
TIME: 18:32:15

Input Set: **I235416A.RAW**

P#10

This Raw Listing contains the General Information Section and up to first 5 pages.

<110> APPLICANT: Sakowicz, Roman
Goldstein, Lawrence S. B.
The Regents of the University of California
<120> TITLE OF INVENTION: Identification and Expression of a Novel Kinesin Motor Protein
<130> FILE REFERENCE: 18557C-000710US
<140> CURRENT APPLICATION NUMBER: US/09/235,416A
<141> CURRENT FILING DATE: 1999-01-22
<150> EARLIER APPLICATION NUMBER: WO PCT/US99/01355
<151> EARLIER FILING DATE: 1999-01-22
<150> EARLIER APPLICATION NUMBER: US 60/072,361
<151> EARLIER FILING DATE: 1998-01-23
<160> NUMBER OF SEQ ID NOS: 7
<170> SOFTWARE: PatentIn Ver. 2.0
<210> SEQ ID NO 1
<211> LENGTH: 784
<212> TYPE: PRT
<213> ORGANISM: Thermomyces lanuginosus
<220> FEATURE:
<223> OTHER INFORMATION: TL-gamma ATP-dependent plus end-directed microtubule motor protein
<220> FEATURE:
<221> NAME/KEY: DOMAIN
<222> LOCATION: (1)..(357)
<223> OTHER INFORMATION: kinesin-like microtubule motor domain
<220> FEATURE:
<221> NAME/KEY: DOMAIN
<222> LOCATION: (358)..(442)
<223> OTHER INFORMATION: neck domain links motor domain to stalk domain
<220> FEATURE:
<221> NAME/KEY: DOMAIN
<222> LOCATION: (443)..(601)
<223> OTHER INFORMATION: stalk domain, unc-104 family domain
<220> FEATURE:
<221> NAME/KEY: DOMAIN
<222> LOCATION: (602)..(784)
<223> OTHER INFORMATION: tail domain
<400> SEQUENCE: 1

```
Met Ser Gly Gly Gly Asn Ile Lys Val Val Val Arg Val Arg Pro Phe
  1               5                  10                  15
Asn Ala Arg Glu Ile Asp Arg Gly Ala Lys Cys Ile Val Arg Met Glu
             20                  25                  30
Gly Asn Gln Thr Ile Leu Thr Pro Pro Pro Gly Ala Glu Glu Lys Ala
         35                  40                  45
```

```
Arg Lys Ser Gly Lys Thr Ile Met Asp Gly Pro Lys Ala Phe Ala Phe
     50                  55                  60
Asp Arg Ser Tyr Trp Ser Phe Asp Lys Asn Ala Pro Asn Tyr Ala Arg
 65                  70                  75                  80
Gln Glu Asp Leu Phe Gln Asp Leu Gly Val Pro Leu Leu Asp Asn Ala
                 85                  90                  95
Phe Lys Gly Tyr Asn Asn Cys Ile Phe Ala Tyr Gly Gln Thr Gly Ser
            100                 105                 110
Gly Lys Ser Tyr Ser Met Met Gly Tyr Gly Lys Glu His Gly Val Ile
        115                 120                 125
Pro Arg Ile Cys Gln Asp Met Phe Arg Arg Ile Asn Glu Leu Gln Lys
    130                 135                 140
Asp Lys Asn Leu Thr Cys Thr Val Glu Val Ser Tyr Leu Glu Ile Tyr
145                 150                 155                 160
Asn Glu Arg Val Arg Asp Leu Leu Asn Pro Ser Thr Lys Gly Asn Leu
                165                 170                 175
Lys Val Arg Glu His Pro Ser Thr Gly Pro Tyr Val Glu Asp Leu Ala
            180                 185                 190
Lys Leu Val Val Arg Ser Phe Gln Glu Ile Glu Asn Leu Met Asp Glu
        195                 200                 205
Gly Asn Lys Ala Arg Thr Val Ala Ala Thr Asn Met Asn Glu Thr Ser
    210                 215                 220
Ser Arg Ser His Ala Val Phe Thr Leu Thr Leu Thr Gln Lys Trp His
225                 230                 235                 240
Asp Glu Glu Thr Lys Met Asp Thr Glu Lys Val Ala Lys Ile Ser Leu
                245                 250                 255
Val Asp Leu Ala Gly Ser Glu Arg Ala Thr Ser Thr Gly Ala Thr Gly
            260                 265                 270
Ala Arg Leu Lys Glu Gly Ala Glu Ile Asn Arg Ser Leu Ser Thr Leu
        275                 280                 285
Gly Arg Val Ile Ala Ala Leu Ala Asp Met Ser Ser Gly Lys Gln Lys
    290                 295                 300
Lys Asn Gln Leu Val Pro Tyr Arg Asp Ser Val Leu Thr Trp Leu Leu
305                 310                 315                 320
Lys Asp Ser Leu Gly Gly Asn Ser Met Thr Ala Met Ile Ala Ala Ile
                325                 330                 335
Ser Pro Ala Asp Ile Asn Phe Glu Glu Thr Leu Ser Thr Leu Arg Tyr
            340                 345                 350
Ala Asp Ser Ala Lys Arg Ile Lys Asn His Ala Val Val Asn Glu Asp
        355                 360                 365
Pro Asn Ala Arg Met Ile Arg Glu Leu Lys Glu Glu Leu Ala Gln Leu
    370                 375                 380
Arg Ser Lys Leu Gln Ser Ser Gly Gly Gly Gly Gly Gly Ala Gly Gly
385                 390                 395                 400
Ser Gly Gly Pro Val Glu Glu Ser Tyr Pro Pro Asp Thr Pro Leu Glu
                405                 410                 415
Lys Gln Ile Val Ser Ile Gln Gln Pro Asp Ala Thr Val Lys Lys Met
            420                 425                 430
Ser Lys Ala Glu Ile Val Glu Gln Leu Asn Gln Ser Glu Lys Leu Tyr
        435                 440                 445
```

```
Arg Asp Leu Asn Gln Thr Trp Glu Glu Lys Leu Ala Lys Thr Glu Glu
    450                 455                 460
Ile His Lys Glu Arg Glu Ala Ala Leu Glu Glu Leu Gly Ile Ser Ile
465                 470                 475                 480
Glu Lys Gly Phe Val Gly Pro Tyr His Ser Lys Glu Met Pro His Leu
                485                 490                 495
Val Asn Leu Ser Asp Asp Pro Leu Leu Ala Glu Cys Leu Val Tyr Asn
            500                 505                 510
Ile Lys Pro Gly Gln Thr Arg Val Gly Asn Val Asn Gln Asp Thr Gln
        515                 520                 525
Ala Glu Ile Arg Leu Asn Gly Ser Lys Ile Leu Lys Glu His Cys Thr
    530                 535                 540
Phe Glu Asn Val Asp Asn Val Val Thr Ile Val Pro Asn Glu Lys Ala
545                 550                 555                 560
Ala Val Met Val Asn Gly Val Arg Ile Asp Lys Pro Thr Arg Leu Arg
                565                 570                 575
Ser Gly Tyr Arg Ile Ile Leu Gly Asp Phe His Ile Phe Arg Phe Asn
            580                 585                 590
His Pro Glu Glu Ala Arg Ala Glu Arg Gln Glu Gln Ser Leu Leu Arg
        595                 600                 605
His Ser Val Thr Asn Ser Gln Leu Gly Ser Pro Ala Pro Gly Arg His
    610                 615                 620
Asp Arg Thr Leu Ser Lys Ala Gly Ser Asp Ala Asp Gly Asp Ser Arg
625                 630                 635                 640
Ser Asp Ser Pro Leu Pro His Phe Arg Gly Lys Asp Ser Asp Trp Phe
                645                 650                 655
Tyr Ala Arg Arg Glu Ala Ala Ser Ala Ile Leu Gly Leu Asp Gln Lys
            660                 665                 670
Ile Ser His Leu Thr Asp Asp Glu Leu Asp Ala Leu Phe Asp Asp Val
        675                 680                 685
Gln Lys Ala Arg Ala Val Arg Arg Gly Leu Val Glu Asp Asn Glu Asp
    690                 695                 700
Ser Asp Ser Gln Ser Ser Phe Pro Val Arg Asp Lys Tyr Met Ser Asn
705                 710                 715                 720
Gly Thr Ile Asp Asn Phe Ser Leu Asp Thr Ala Ile Thr Met Pro Gly
                725                 730                 735
Thr Pro Arg Ser Asp Asp Asp Gly Asp Ala Leu Phe Phe Gly Asp Lys
            740                 745                 750
Lys Ser Lys Gln Asp Ala Ser Asn Val Asp Val Glu Glu Leu Arg Gln
        755                 760                 765
Gln Gln Ala Gln Met Glu Glu Ala Leu Lys Thr Ala Lys Gln Glu Phe
    770                 775                 780
<210> SEQ ID NO 2
<211> LENGTH: 2352
<212> TYPE: DNA
<213> ORGANISM: Thermomyces lanuginosus
<220> FEATURE:
<223> OTHER INFORMATION: TL-gamma ATP-dependent plus end-directed
      microtubule motor protein
<400> SEQUENCE: 2
```

atgtcgggcg gtggaaatat caaggtggtg gtgcgggtac gcccgttcaa cgcccgagaa 60
atcgaccgtg gcgcaaaatg tattgtgcgg atggaaggaa atcaaaccat cctcaccct 120
cctccgggtg ccgaagagaa ggcgcgtaaa agtggcaaaa ctattatgga tggcccgaag 180
gcatttgcgt tcgatcggtc gtattggtcc tttgacaaga atgctcccaa ctatgcgaga 240
caggaagacc tattccaaga tctcggagtc ccgcttctgg ataatgcatt caagggttat 300
aacaattgta tcttcgccta cggtcagacc ggttcgggca agtcctattc aatgatgggc 360
tatggcaagg agcatggcgt gatcccgcgg atttgccagg acatgttccg gcgtattaat 420
gaactgcaga aggacaagaa cctcacttgc accgtcgaag tttcgtactt ggaaatttac 480
aatgaacgag tgcgagactt gctgaatccg.tcgacaaagg ggaatctcaa ggtccgagaa 540
cacccgtcga ccggccccta cgtggaggac ttggcgaagc tggtcgtgcg atcattccaa 600
gaaatcgaaa atctcatgga tgagggcaac aaagccagaa cggttgccgc cacaaacatg 660
aacgagacat ccagtcgatc ccacgccgtc ttcactttga ccttgacgca aaagtggcat 720
gatgaagaga ccaaaatgga cacagagaag gttgcgaaga tcagtctggt agatttggcg 780
ggttctgagc gagcaacgtc caccggagct actggagcgc gactgaagga gggtgcagag 840
atcaaccgct cactttcgac cctaggtcgt gtgattgcag cgctagcgga tatgtcgtcg 900
ggaaaacaga agaagaatca gttagtacct taccgagatt cggtactgac gtggcttctg 960
aaggactcct tgggaggcaa ctcgatgacc gccatgattg ccgccatttc gcctgctgat 1020
attaactttg aagagactct cagtaccctt cgatatgcgg actctgcgaa gcgaatcaag 1080
aaccacgcag tggtcaatga agacccgaac gcgcggatga tccgcgagtt gaaggaggaa 1140
ctcgcgcagc tgaggagcaa actccagagc agtggtggag gtggaggtgg tgcaggaggt 1200
tctggcgggc cagtggagga atcgtacccg cccgacacgc cgctcgagaa gcaaatcgtg 1260
tcgattcagc agccggatgc gacagtcaag aaaatgagca aggcagaaat cgtggagcaa 1320
ctgaaccaga gtgagaagct ctatcgggat ctcaatcaga cctgggaaga gaagctggcc 1380
aagaccgagg aaatccacaa ggaacgagaa gcggcgctcg aggagctggg tatcagcatc 1440
gaaaagggct ttgttggccc ttaccactcc aaagaaatgc cacatctagt caacttgagc 1500
gatgatcctc ttctggctga gtgtcttgtc tacaacatca agcccgggca gacaagggtt 1560
ggaaacgtca accaagatac acaagcggaa attcgtctga acggttcgaa gatcctgaaa 1620
gaacactgta cgtttgaaaa tgtggacaac gttgtgacca tcgtgccaaa cgagaaggct 1680
gctgtcatgg tgaacggcgt gcgaatcgac aagcctactc gcctccgcag cggctacagg 1740
atcatcctgg gcgatttcca catttttcga ttcaaccatc cggaagaagc tcgtgcggaa 1800
cggcaagaac aatccttgct tcgccattct gtcaccaaca gtcagttggg ttcgcctgct 1860
ccaggccgtc acgaccggac actgagcaag gcgggttcgg atgcggacgg cgattctcgc 1920
tcagattctc ctttgccgca ctttcgtgga aaggatagcg actggttcta tgctcgcagg 1980
gaagctgcta gcgcgatcct agggttggat cagaagatct ctcatctgac agatgacgag 2040
ttggatgcat tatttgacga tgttcagaaa gcgcgggcag ttcgtcgtgg gctggtcgaa 2100
gacaacgaag atagcgattc gcagagttcg tttccggtcc gtgacaaata catgtccaat 2160
ggaaccattg ataatttctc gctcgatacc gccattacta tgccgggtac ccctcgtagt 2220
gatgacgacg gtgacgcgct gtttttggt gataagaagt cgaaacagga tgcgtctaat 2280
gttgatgttg aggagttgcg tcaacagcag gctcagatgg aagaagccct gaaaacagcg 2340
aagcaggaat tc 2352

<210> SEQ ID NO 3
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:primer
<400> SEQUENCE: 3
atgtcgggcg gtggaaatat c 21

<210> SEQ ID NO 4
<211> LENGTH: 23

```
195  <212> TYPE: DNA
196  <213> ORGANISM: Artificial Sequence
197  <220> FEATURE:
198  <223> OTHER INFORMATION: Description of Artificial Sequence:primer
199  <400> SEQUENCE: 4
200        gaattcctgc ttcgctgttt tca                                    23
201  <210> SEQ ID NO 5
202  <211> LENGTH: 30
203  <212> TYPE: DNA
204  <213> ORGANISM: Artificial Sequence
205  <220> FEATURE:
206  <223> OTHER INFORMATION: Description of Artificial Sequence:degenerate
207        forward primer
208  <220> FEATURE:
209  <221> NAME/KEY: modified_base
210  <222> LOCATION: (25)
211  <223> OTHER INFORMATION: n = a, c, g or t
212  <400> SEQUENCE: 5
W--> 213  gcgcggatcc atyttygcht ayggncarac                              30
214  <210> SEQ ID NO 6
215  <211> LENGTH: 30
216  <212> TYPE: DNA
217  <213> ORGANISM: Artificial Sequence
218  <220> FEATURE:
219  <223> OTHER INFORMATION: Description of Artificial Sequence:degenerate
220        reverse primer
221  <220> FEATURE:
222  <221> NAME/KEY: modified_base
223  <222> LOCATION: (16)
224  <223> OTHER INFORMATION: n = a, c, g or t
225  <220> FEATURE:
226  <221> NAME/KEY: modified_base
227  <222> LOCATION: (28)
228  <223> OTHER INFORMATION: n = a, c, g or t
229  <400> SEQUENCE: 6
W--> 230  gcgcgaattc tcdganccdg cvarrtcnac                              30
231  <210> SEQ ID NO 7
232  <211> LENGTH: 30
233  <212> TYPE: DNA
234  <213> ORGANISM: Artificial Sequence
235  <220> FEATURE:
236  <223> OTHER INFORMATION: Description of Artificial Sequence:degenerate
237        reverse primer
238  <220> FEATURE:
239  <221> NAME/KEY: modified_base
240  <222> LOCATION: (16)
241  <223> OTHER INFORMATION: n = a, c, g or t
242  <220> FEATURE:
243  <221> NAME/KEY: modified_base
244  <222> LOCATION: (28)
```

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

FYI

VERIFICATION SUMMARY
PATENT APPLICATION **US/09/235,416A**

Input Set: **I235416A.RAW**

Line	?	Error/Warning	Original Text
213	W	"N" or "Xaa" used: Feature required	gcgcggatcc atyttygcht ayggncarac
230	W	"N" or "Xaa" used: Feature required	gcgcgaattc tcdganccdg cvarrtcnac
247	W	"N" or "Xaa" used: Feature required	gcgcgaattc tcdctnccdg cvarrtcnac